https://dx.doi.org/10.22161/ijaers.4.1.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

A Competent Approach for Type of Phishing Attack Detection Using Multi-Layer Neural Network

Bhawana Goyal[1], Meenakshi Bansal[2]

[1]M. Tech Scholar (Computer Engineering), Guru Kashi Campus, Punjabi University, Patiala, India
[2]Asst. Proff, Guru Kashi Campus, Punjabi University, Patiala, India

Abstract— *With the enlargement of contemporary technologies and the large-scale global computer networks web-attacks are escalating because of emergent curiosity of people and lawful institutions towards internet. Phishing is one of web-attack carried out by attacker using both social and technical engineering. Generally on web more attacks are launched every month with seek of crafting web addict to consider that they are contacting with a legalized entity for the intention of embezzle identity information, logon records and account details. The phishing attack detection and classification methods are utilized for the prevention and in-depth analysis of the attacks. In this paper, the proposed model has been designed with the multi-directional feature analysis along with the Back-Propagation Probabilistic neural network (BP-PNN) classification. The proposed model has performed better in the terms of the accuracy in all of the domains based upon the attack detection and classification.*

Keywords—Phishing detection, Neural Network, Probabilistic Classification, Phishing Attack Classification.

I. INTRODUCTION

Phishing is fraudulent process carried out by phisher to cheat people by creating spurious websites that have same manifestation to valid one to mislead individual's personal information [2]. Primarily focus of phisher is to amass monetary information, economic fatalities, identity hiding, loss in ecommerce and e-banking. Presently, Researchers have studied that phishing attacks have broadened beyond emails & append instant messaging, social networking sites, multiplayer games etc. Mostly used phishing attack is spear phishing in which attacker sends fake emails to specific target people rather than sending mass emails without knowing who the victim will be. For that reason, now Anti-phishing is multifaceted, rigid and critical disaster in recent humanity. Due to increasing online business it's very important to grasp this problem.

A. Types of Phishing Attacks

- Spear phishing - In this type of phishing [5] attacker have definite target of individuals and companies from whom he/she possibly will collect delicate information. It is most successful on web having 91% of phishing attacks.
- Whaling - Has tricky messages intended to look similar as received from actual business organization. Here phisher creates authoritatively looking FBI written order and ensures executive at upper level will click link and launch unusual programming.
- Clone phishing - In this beneficiary addresses are taken and used to make a practically indistinguishable or we can say duplicate (clone) email.
- Link manipulation - Traps used in this type of phishing is wrong spelling of URL and also sub-domains are utilized.
- Website forgery - Once a sufferer visits the phishing website the trickery isn't over. Scams use JavaScript commands so as to change the address bar. This is done by putting an image of a legitimate universal resource locator over the address bar.
- Phone phishing – Clients get fake [8] messages and phisher guaranteed that these are from bank advising to dial a telephone no. in regards to issues with their accounts.

B. Anti-Phishing Solutions

- Email Signed Digitally
- Online Brand Checking
- Web Browser Extensions
- Browsers Alert To unreliable Websites
- Identify Valid Websites

Also apart from these solutions there are more techniques and approaches proposed by researcher and anti-phishing [6] working groups (APWG) that are set up in various countries which detect and classify these phishing attacks.

https://dx.doi.org/10.22161/ijaers.4.1.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

II. RELATED WORK

Authors proposed an efficient approach for phishing detection using single-layer neural network. In this new phishing websites can be detected & the weights of the heuristic are derived dispassionately [1]. In this paper new neuro-fuzzy model is proposed to detect phishing sites without by means of rule sets. Membership function has been used to calculate the value of heuristics efficiently [2]. Classification of large volume deception sites is very pricey both economically and computationally. To reduce problems like financial and computation that can be solved by distributed cloud environment [3]. Intelligent model for detection of fraud emails and websites by extracting features of emails & url by using preprocessing phase. New premeditated model results of accuracy 98.87% by using random forest algorithm (RFA) having standard dataset [4].Dudhe [5] presented that phishing is attack on web which make use of technical & social engineering to formulate phony websites and messages to fool users on internet to have their personal information. Various approaches have been applied to detect forged websites such as blacklist, whitelist, heuristic-based, machine learning approach. Presently increasing problems for internet users due to attacks on internet by falsified websites and emails which to attain sensitive information [6]. In this neural networks are used to predict the phishy websites [7]. This paper has multilayer neural network minimize error and make higher the performance. Discovered diverse features concerning the attacks of phishing, problems caused by them and projected the problem statement to seek out the supreme result for the dilemma [8]. Google Page Rank based approach was proposed for detection of phishing website. The proposed techniques have four phases. Further the features for phishing sites were selected. To get results classification process is carried out for testing and accuracy [9].In this paper authors used 2889 phishing and legitimate emails for revise, to test and train the classifiers 43 additional features have been used [10]. HTML is broadly use for webpages formation in computer network of Internet and Intranet [11]. Source code of HTML and JavaScript is obtained while webpages load in browser. In this they had presented different algorithms for encryption and decryption of html & JavaScript. Thabtah et al. [12] has predicted the websites phishing detection using neural network trained with back-propagation. It shows high reception ability for fault tolerance, noisy data and other parameters. In this paper approach to classify vehicles based on probabilistic neural network and features are extracted with feature extractor [13].The intellectual Phishing Website Detection System using Fuzzy method for E-Banking that have advantage to facilitate dispensation of hazily distinct variables and their relationship between them [14].Phishing is web attack done by attacker while online transaction or social media to grasp personal details of victim [15]. In this detail information is provided regarding anti-phishing techniques along with their advantage and disadvantages. Sung et al. [16] had projected that how to detect phishing URLs using different approaches designed by organizations. Here comparison is done with previous machine learning approaches with newly proposed real-time application.

III. PROPOSED WORK

Phishing [1] is untruthful process carried out by phisher using both social and technical engineering to swindle people by crafting spurious websites that have same manifestation to legal one to mislead individual's delicate information. From most current statistics phishing is escalating felony on web that includes major loss of capital and perceptive information communal on web. The proposed model has been analyzed under the variety of the experiments and has been specifically designed for the detection and classification of phishing attacks. Some of phishing attacks included in proposed work are Clone phishing, Link manipulation, Phone phishing, Spear phishing, Website Forgery and Whaling method. Here artificial neural network is used as classifier for classification and detection [12] of phishing attacks. Due to non-linear nature of neural networks, they consists large number of processing elements called neurons. To bare the patterns in data neural networks have complex connection between inputs and outputs. The most interesting feature of neural network is the possibility of learning. Also fuzzy rules have been estimated here to remove crispy values.

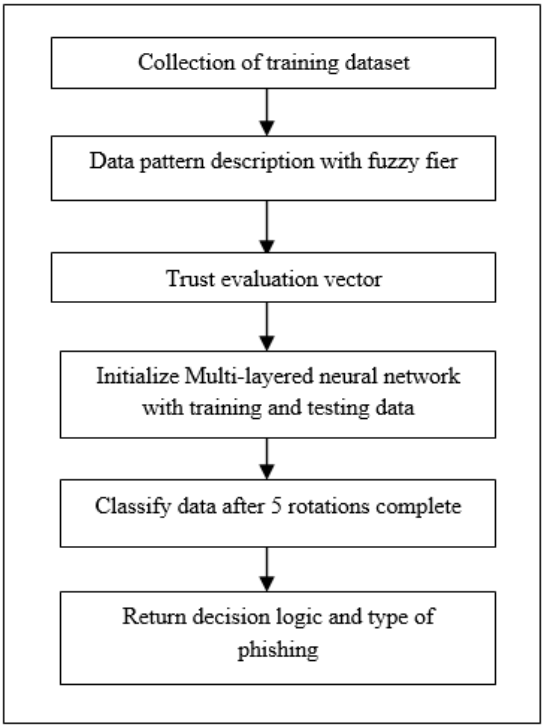

Fig.1: Flowchart of Methodology

https://dx.doi.org/10.22161/ijaers.4.1.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

In this paper, a competent approach for detecting type of phishing attack based on multi-layer neural network is proposed having output on the basis of similarity.

A. Artificial Neural Networks

An Artificial neural network (ANN) repeatedly now called "neural network" (NN) is a statistical model based on genetic neural networks [17] that look a lot like the brain. Due to non-linear nature of neural networks, they consists large number of processing elements called neurons. To uncover the patterns in data neural networks have complex connection between inputs and outputs. Neural network as classifier works in similar way humans' processes information. The interneuron association strength called synaptic weights mount up knowledge that the neural networks require from a learning process. Neural network have large amount of nerve cells which are extremely simple that work similar and have the ability to learn. These artificial neural networks (ANN) fluctuate from each other with one or two layers of single route logic, to complicated multi-input with many directional feedback loops and inputs.

From above observation we can say that phishing is classification problem that can be solved by different classifiers [10] in data mining that are Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Neural Networks (NN).In this paper to propose our model Neural Networks (NN) are used because it has quite a lot of advantages like high precision, noise lenience, relieve of maintenance, independence from earlier assumptions. Neural networks are worn to predict type of phishing attack with the help of its learning and generalization aspects.

B. System Model Design

In previous work the model was designed [1] in which single-layer neural network was intended with 11,660 training dataset including phishy & legal sites and 2 testing dataset having 5000 phishing sites or 5000 legitimate sites over which testing is done. In our proposed model we had used multi-layer neural network with 32000*2 phishy dataset and there are 12 samples as testing dataset. Here fuzzy [14] fier is also used for reducing data or we can say feature reduction is done. We have all phishy patterns because main objective is to detect and classify type of phishing attack.

1. Phase I - In this 20 main features are extracted some of them are Trust factor, No. of nodes, each node priority, each node trust factor, ranking, patterns etc.
2. Phase II - Then trust evaluation vector calculates the trust value on the basis of prominent pattern selection.
3. Phase III - Here in this phase Multi-layered neural network is trained with training data to calculate value of output node.
4. Phase IV – Classify phishy data after 5 rotations compute.
5. Phase V – At last decision logic is generated after neural network computes 30 times and type of phishing attack will be result.

C. Classification using Back Propagation Probabilistic Neural Network (BP-PNN)

A Probabilistic neural network (PNN) [13] is neural network which is plagiaristic of Bayesian network and kernel fisher discriminant analysis. It is feed forward neural networks, which has layered structure. In this layered formation one or more processing elements are present at each layer. No additional extensive training is required while training samples are added or removed.

Here due to layers each processing aspect gets input from previous layer or also can get from outside world. A PNN is an accomplishment of numerical algorithm called kernel discriminant analysis where operations are keen on multilayered feed forward neural network with four layers named as:

- Input layer
- Hidden layer
- Summation layer/Pattern layer
- Output layer

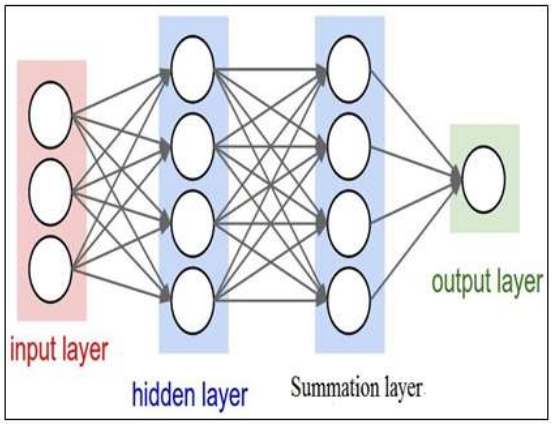

Fig.2: Architecture of probabilistic neural network

In the implementation of phishing pattern discovery algorithm (PPDA) using the multiple input layers initialized with the variety of the features extracted. The phishing pattern discovery module adds the robustness to the proposed classification model to process input data according to the target data. Testing of the input web patterns and their recognization as the phishing patterns is done by using different functions like performance, error evaluation and feature extraction modules [6].

D. Proposed System

For the purpose type of phishing attack classification, proposed system has datasets i.e. training dataset consists type of phishing attacks spear phishing (s) ,whaling (z) ,clone phishing (c) ,phone phishing (p) , website forgery (w) ,link manipulation (l) and testing dataset consist samples

https://dx.doi.org/10.22161/ijaers.4.1.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

(n=12).Also database is created which have multiple data i.e. 32000*2 phishy data .Then additional neural network is initialized for processing which evaluates 5 rotations on each of attack i.e. NN will compute 30 times. From here we can examine that NN [7] will evaluate result after computing whole sample. The end result will be having parameter i.e. similarity (maximum similar results).Moreover graphs are also generates in pyramid and triangular shape by neural network meanwhile it is trained.. We compute NN 5 times so that it can converge and accuracy can be obtained.

Steps of Training Algorithm [1] [2] :-

1. Firstly weights of heuristics are initialized and collect training dataset.
2. Then neural network structured access these weights as input and in forward direction these are transmitted to whole network until it appears at output layer. Afterward equation is used to calculate input for output layer:

$$O_I = \sum_{i=1}^{n} W_i \times I_i \quad \ldots\ldots\ldots\ldots\ldots (1)$$

Here in this equation O_I is input value for output layer, W_i are the weights initialized for ith input node and I_i is the input value of the ith node. Now here output value of output node is calculated by equation below:

$$O_o = \frac{1}{1+ e^{-o_I}} \quad \ldots\ldots\ldots (2)$$

3. To calculate the error we deduct output value of output node from real output value.

$$err = T - O_o \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

where T is real output.

4. To train neural network supervised learning is employed, in this back-propagation probabilistic neural network algorithm is used. In this below equation is used to adjust weights:

$$W_i = W_i \times R \times err \times O \quad \ldots\ldots\ldots (4)$$

Where W_i is weight of input it node and R is learning rate here.

5. When neural network stops then forward procedure begins again until the error is minimized between valid output and predicted output.

IV. RESULTS AND DISCUSSION

In this network, there is ten input layer with seventy four neurons, ten hidden layers and an output layer having one neuron. Features which have been previously extracted are saved in dataset. There are 20 numbers of features that are extracted. These features are supplied to input layer of Feed Forward NN (FFNN). Data will be processed for classification in hidden layers. Output layer will provide the result phishing sample and identify it according to the target value. Test the performance of overall system on the basis of accuracy, precision, recall and elapsed time.

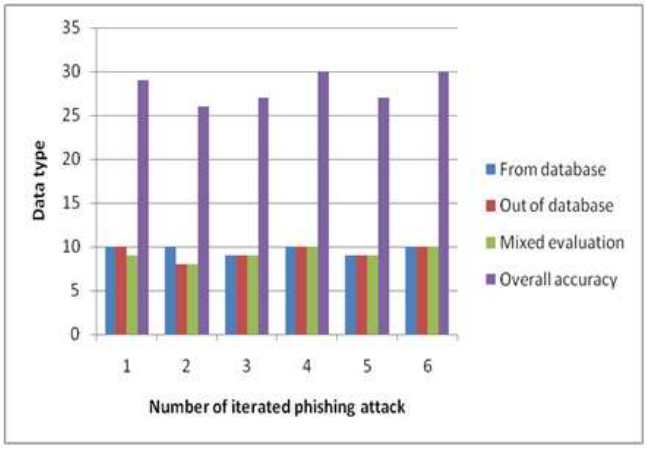

Fig.3: Accuracy analysis of iterated phishing attack

Figure.3 portrays accuracy analysis of iterated phishing attack recognition system that is detected in proposed model. Here in column graph number of phishing attack is on x-axis and data type is on y-axis. Classification method can be used to classify dataset into one of seeded set of classes or groups. Training dataset consist of types of phishing attacks i.e. on x-axis: spear phishing (s) ,whaling (z) ,clone phishing (c) ,phone phishing (p) , website forgery (w) ,link manipulation (l). Sequence followed to detect accuracy of phishing attack is clpswz.

Elapsed Time: is defined as measurement of time completing an activity, job or task. In other words, it is defined as difference between finishing time and starting time of the neural network.

Elapsed Time =Finishing Time –Starting Time.

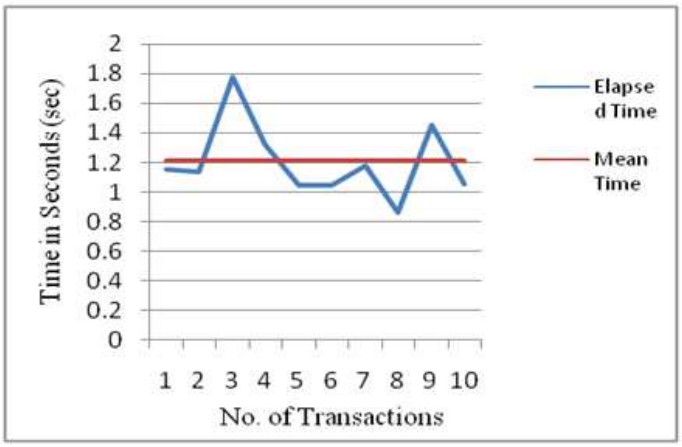

Fig.4: Elapsed time graph of neural network

Figure 4 shows elapsed time graph of neural network. X-axes represent no of transactions and y-axes represent time in seconds. Some point of elapsed time is greater than average time of neural network and some point of elapsed

https://dx.doi.org/10.22161/ijaers.4.1.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

time is below the average time of neural network. This graph shows two lines blue and red. Blue line represent elapsed time of neural network .Red line represents mean time (average time) of elapsed time of the neural network.

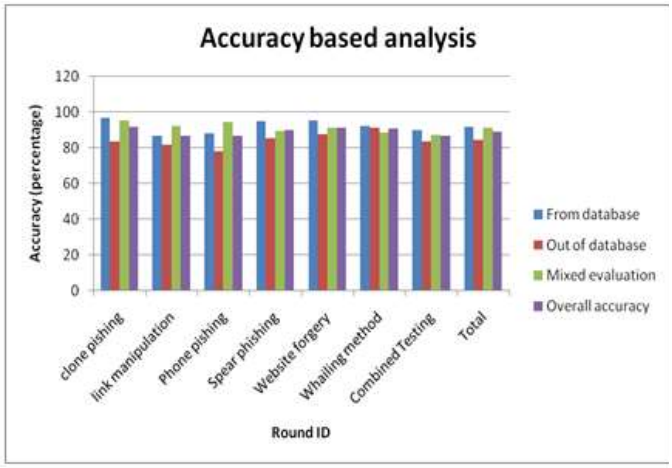

Fig.5: Accuracy analysis of system in percentage

Figure.5 above illustrates accuracy analysis of system in percentage. Column graph demonstrates two axis where x-axis correspond to different type of phishing attack, combinational testing, total of them and y-axis correspond to accuracy in (percentage) among various data type i.e. From database, Out of database, Mixed evaluation, Overall accuracy.

Also in figure.5 same sequence is followed that is clpswz i.e. shown above and the proposed model is sculpt for classification of phishy data and then detect which type of phishing attack is it.

V. CONCLUSION AND FUTURE WORK

This research work proposed an innovative technique to detect and classify type of phishing attack. In this new proposed technique, the system sculpt is built to detect phishing attack using multi-layer neural network. In consequence, the boosted model formation is necessitating for steady revolutionize and buffet these changes in websites. As discussed above phishing is fraud carried out by invader on web is swindling to divulge individual's credentials like account details, user name, passwords etc. It's critical impasse to be foretell to defeat identity stealing and economic losses which are its commencement. The multi-layer neural network Back-propagation probabilistic neural network (BP-PNN) is experimented with large database i.e. 32000*2 phishy data. Neural network is trained with 6 types of phishing attacks and 12 testing samples for classification of phishy data. So from this we believe that as compared to existing system tools the proposed system has better performance and less error rates.

In future the projected multi-layer neural network will be enlarged to improve the detection ratio. Also more classifiers like SVM, KNN, DT, NB etc. can be used for classification and further comparison can be done to get best results among them. Additional features can be used to detect newly created phishing attacks by invader on web.

REFERENCES

[1] Nguyen, L.A.T., To, B.L., Nguyen, H.K., and Nguyen, M.H. (2014, October).An efficient approach for phishing detection using single-layer neural networks. In International Conference on Advanced Technologies for Communications (ATC 2014), pp.435-440, IEEE.

[2] Nguyen, L.A.T., To, B.L., and Nguyen, H.K. (2015, December). An Efficient Approach for Phishing Detection Using Neuro-Fuzzy Model. Journal of Automation and Control Engineering, Vol.3 (6), pp. 519-525.

[3] Shrestha, N., Kharel, R.K., Britt, J., and Hasan, R. (2015, June). High-performance Classification of Phishing URLs using a Multi-modal Approach with MapReduce. In 2015 IEEE World Congress on Services, pp. 206-212.

[4] Smadi,S.,Aslam,N.,Zhang,L.,Alasem,R.,andHossain, M.A.(2015,December).Detection of Phishing Emails using Data Mining Algorithms. In 2015 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), pp.1-8, IEEE.

[5] Dudhe, P.D., and Ramteke, P.L. (2015, February). A review on phishing detection approaches. Journal of Computer Science and Mobile Computing, Vol.4 (2), pp.166 – 170.

[6] Sathya, R., Vijayaraj, J., and Purushotaman, P. (2015, March). Techniques to prevent the users from phishing attacks. Journal of Modern Trends in Engneering and Research, Vol.2 (3), pp.251-255.

[7] Martin,A.,Anutthamaa,N.,Sathyavathy,M.,Francois,M. M.S., and Venkatesan,D.V.P.(2011, September). A Framework for Predicting Phishing Websites Using Neural Networks. Journal of Computer Science Issues, Vol.8 (2), pp.330-336.

[8] Sahu, K.R., and Dubey, J. (2014, January). A Survey on Phishing Attacks. Journal of Computer Applications, Vol. 88(10), pp.42-45.

[9] Sunil, A.N.V., and Sardinia, A. (2012). A Reputation Based Detection Technique to Web Spam. In Department of Electronics and Computer Engineering, pp.566-572, Elsevier.

[10] Abu-Nimeh, S., Nappa, D., Wang, X., and Nair, S. (2007, October). A Comparison of Machine Learning Techniques for Phishing Detection. In Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit pp. 60-69, (ACM).

https://dx.doi.org/10.22161/ijaers.4.1.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

[11] Zhongying, B., and Jianchen,Q. (2009,August).Webpage Encryption Based on Polymorphic JavaScript Algorithm. In Information Assurance and Security, IAS'09, Fifth International Conference, Vol.1, pp.327-330, IEEE.

[12] Mohammad,R.M.,Thabtah,F., and McCluskey,L.(2013,Janruary). Predicting Phishing Websites using Neural Network trained with Back-Propagation. In Proceedings of International Conference on Artificial Intelligence p.1, (ICAI).

[13] Mishra,M.,Jena,A.R., and Das,R.(2013,July). A Probabilistic Neural Network Approach For Classification of Vehicle , Journal of Application or Innovation in Engineering & Management, Vol.2(7) , pp.367-371.

[14] Gandhi, R., and Backiyalakshmi, R. (2014, October). Intelligent Phishing Website Detection System using Fuzzy Technique for E-Banking.Vol.24 (82), pp.33-40.

[15] Hussain, M.R., and Srivastava, G. (2014, May). Phishing a Growing Scam- a Review Paper. Journal of Advance Research in Science and Engineering, Vol.3 (5), pp.85-93.

[16] Basnet, R.B., Sung, A.H., and Liu, Q.(2014,June). Learning to detect phishing URLs. Journal of Research in Engineering and Technology, Vol.3 (6), pp.11-24.

[17] Singh, Y., and Chauhan, A.S. (2009). Neural Networks In Data Mining. Journal of Theoretical and Applied Information Technology, Vol.5 (6), pp. 37-42.